This document is published in:

Cabestany, Joan; et al. (eds.) (2011) *Advances in Computational lntelligence: 11th lnternational WorkConference on Artificial Neural Networks, IWANN 2011, Torremolinos-Málaga, Spain, June 8-10, 2011, Proceedings, Part II.* (Lecture Notes in Computer Science, 6692) Springer, pp. 645-652.
DOI: http://dx.doi.org/10.1007/978-3-642-21498-1_81

Group Behavior Recognition in Context-Aware Systems

Alberto Pozo, Jesús Gracía, Miguel A. Patricio, and José M. Molina

GIAA, Carlos III University, Spain
{alberto.pozo,jesus.garcia,miguelangel.patricio,
josemanuel.molina}@uc3m.es

Abstract. In most of the domains of the context-aware system the user make up a group and their behavior could be research with group behavior recognition techniques. Our approach try to take advantage of the context information to understand the user's behaviors like a group, and this information could be useful for other system to beat to the users. For this purpose it is exposed a new representation that concentrates all necessary information concerning relations two to two present in the group, and the semantics of the different groups formed by individuals and training (or structure) of each one of them.

Keywords: Group behavior recognition, context-aware systems, activity representation, computer vision.

1 Introduction

Pervasive computing is one of the most active fields of research. This discipline need the context information to beat user's behaviors, it need to understand and predict their context and the user's behavior.

For this reason the group behavior recognition techniques could collaborate in the domains where the users interact between them. Our approach tries to use the context information (especially the users position, but also others) to understand the users behavior.

Human activity analysis and behavior recognition has received an enormous attention in the last two decades of computer vision community. A significant amount of research has addressed to behavior recognition of one element in the scene.

Instead of modeling the activities of one single element, group behavior recognition deals with multiple objects and/or people, who are part of groups.

In behavior recognition there are two distinct philosophies for modeling a group; the group could be dealt as a single group (crowd) or as a composition of individuals with some shared objectives. On the pervasive computing domains the users are clearly differentiable, so the crowd perspective could not be appropriate. For this reason in this paper we focus the investigation in the second philosophy, where take place many distinguishable agents and how the context information could be useful for the task.

The present paper shows a new representation of the possible variables existed in the problem. This had been designed to put in order briefly the essential information of the system.

With the aim of achieve our project, it will rely on three levels of abstraction.

Firstly, a matrix will be established with the information of each binary relationship between any individual of the system. This matrix store one vector for each relationship with the features selected to the problem domain. It is important to emphasize that in many case the features selected will include the relative position vector.

In these terms and conditions, for each frame in the video, the outstanding information has being kept, inclusive the geometrical information.

Once being contained all the important information, the process continues in a second abstraction's level where the challenge is capturing the logical information implicated between the communication of individual and groups. For this reason it is necessary to make different combinations for representing every group of the system.

It is a relevant detail to remark that each individual can belong to a several groups at the same time, and the groups have the possibility to incorporate an undefined number of other groups or individual.

In the third level, a new representation is created to reduce the dimension of the problem. One of the important key in this type of domains is that the number of relations between the elements of the scene growth exponential in relation with the number of elements. For this reason, a new representation is created to save the essential information of each group without saving all the relations between each element. Instead of save all the possible edges in a graph, this approach only save important graph that can provide all the important information wasting less space.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes the problem. Section 4 introduces our description. Conclusions are drawn in section 5.

2 Related Work

Context-aware system have became in very important field of research in the last years especially because the appearance of the handheld device. These devices need to know the context, even to predict their context.

In papers (1) or (2) we can see the need for recognize the device's context and understand it. There are a lot of research about the sensors and the way of store all the information. In (3) shows the important of understand the context and what requirements needs to have a context-aware system.

Often the context-aware systems are implemented on handheld device, wearable computers, etc. and their context depends on the user's behavior. So if we need to predict the device's context, we have to predict the users' behavior.

The users of a context-aware system rarely are isolated, so their behavior depends on the group's interactions, nearby users, etc.

Despite the fact that there is plenty of work on single object activities, (4) the field of group activities recognition is relatively unexplored.

Group behavior recognition is a complex task that can be observed under different points of view.

There are two big families of approaches, one logical and one geometrical.

The logical approaches (5) are focused in construct a context-free grammar to describe the group activity based in the individual activities of the group's member.

The main characteristic of this point of view is the important of the first level, the features extraction. They need a previous system that recognizes the activity of each element of the scene.

The geometrical approaches (6), (7) have a different point of view. The features extracted in this case are based on the coordinates of the elements of the scene.

This approaches use to have higher computational complexity and the number of the elements in the scene could become very important.

There are also approaches than combine both perspective, like (8) whose work recognize key elements of a sport (basketball) using the positions of the players. This approach needs to identify the key elements of the domain dealt, and these key elements could be different in many different situations.

One more general approach could be read in (9) where the trajectories of the players (in a Robocup mach) are coded to create set of patterns that identify each type of action.

3 Group Behavior Recognition in Context-Aware Systems

In pervasive computing the context is all the elements (and their relationships) that surround the system. These elements could provide useful information to the system, or could be necessary to predict their state in the nearby future to provide a good service to the user.

Our approach tries to use the context information (especially the users position, but also others) to understand the users behavior.

Group behavior recognition is composed by two steps: in the first one the features of the system should be extracted, and in the second one the features are used to recognize the behavior.

Handheld device, wearable computers, etc. usually have a lot of sensor that could provide information like position, orientation, loud, bright, etc. so in this paper we are going to focus in the second step, we use the extracted features by the devices to make the inference of the behavior.

3.1 General Description

In a general scene there is one area composed of many sub-areas and a number of groups that consist of some elements which could be users or object.

In a group, one element could be related with whatever other element of the group. Each element of the system and each relationship have a set of features (like positioning, color, shape, etc.) The features could suffer changes in time.

Each element of the system should belong to a group, and could belong to many groups at the same time.

It is important to emphasize that any element of the system must be in a group, so there are not isolated elements.

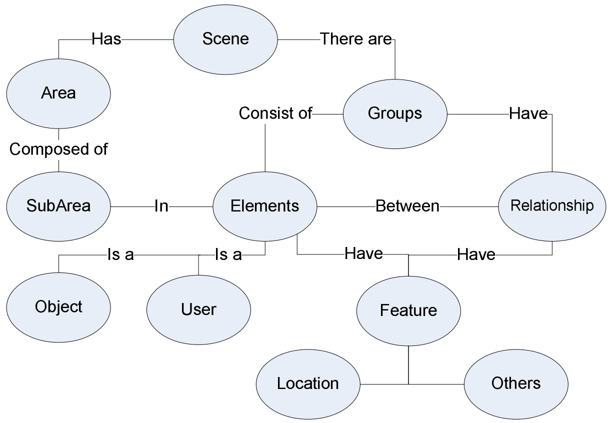

Fig. 1. General scene

3.2 Problem Description

Some of the general axioms of the problem describe above have been eliminated for more practical approach of the problem.

In our approach there is one sequence composed by a number of T instants, where are included a number of N elements (this number cannot change in time). The elements of the scene (all of them users) are distributed in a number of G groups, and each group is represented by a graph. (Group could be composed by two or more elements, and one element could be part of one or more groups).

Each node constitutes one user of the group and each edge constitutes one relationship, graphs are represented by his edges.

For each element and each edge we have a vector of features like positioning, that which is expressed by free vector in polar coordinate system.

Fig. 2 shows a scene with six elements conforming three groups. The definition of the groups is the semantic representation of the relations between the elements of the system.

The features selected to describe the elements of the scene will depend on the problem domain, and it will include its coordinates (in polar coordinate system or spherical coordinate system in case of 3-D positioning) and the coordinates of the free vectors that represents the edges of the graphs.

Fig. 2. Graphic representation of a system with six elements and three groups

For each element and each possible edge we save the features of the vector for each frame of the scene. One feature vector for each element and M free vectors for the edges, where $M = \frac{N \cdot (N-1)}{2}$.

To describe the spatial feature relation between elements i and j in the frame t, there are two coordinates called $|r_{i,j,t}$ and θ_i, with $|r_{i,j,t}| = \sqrt{(x_{j,t} - x_{i,t})^2 + (y_{j,t} - y_{i}}$ and $\theta_{i,j,t} = \tan^{-1}\frac{(y_{j,t}-}{(x_{j,t}-}$.

4 A Structured Representation for the Group Behavior Recognition Issue

Behavior recognition based on the positioning of each element of the group could be helped by the context information (provided by the device's sensors) obtaining better results in any situation.

However, the choice of the features (further away the positioning) it is dependent on the problem domain, so we need to select it in each case.

We propose a structured representation composed by three matrix called R, A and S. The first one save all raw data of the elements (and its relationships) in time (positioning and other dependent on domain features), while the second one represents the information about the semantic of the scene, composed by the number of groups founded and their makeup, and the third one represents the elements features and the important edges structure.

This structured representation contains the information about the features of each element of the scene, the features of the relations between the elements, and the group's structure information.

4.1 Features Vector

As we have wrote above, the structure representation is based on three matrixes. Two of this matrix (R and S) are composed by the features vector. This vector stores the features of one element or one relationship of the system in one definite instant. These features contain the geometrical information and other depending on the domain problem.

Each feature is storage in a natural number between one and eight. The first two features represent the geometrical information, where the first one (d) is the distance between the two elements of the relationship, or between the element and the pole (analogous to the origin of a Cartesian system). And the second one γ is the angle of this distance.

Element positions are calculated by $d_e = \left\lceil \frac{|r_e|\cdot}{|r_{eMax}|} \right.$. Where r_e is the distance between the element and the pole and r_{eMax} is the distance of the most remote element of the pole. Relative distance is calculated by the formula $d = \left\lceil \frac{|r_{ij}|}{|r_{max}|} \right.$. Where r_{ij} is the distance between the elements i and j, and r_{max} is the maximum distance between any elements of the graph. By definition d is a natural number between 1 and 8.

Direction between two elements of the graph (or to positioning one element) is defined by the formula $\gamma = \left\lceil \frac{|\theta_{ij}|}{\pi} \right.$. Where $_{ij}$ is the angle between the elements i and j (or the element and the pole). By definition γ is a natural number between 1 and

8. It is important to perceive that in spite of the graphs are not directed, to construct the reduced graph we have to distinguish between the same directions with different sense. So the possible directions are covered between - π and π radians.

All other features are calculated in the same way, using the max value of the feature, and the result is another natural number between one and eight.

4.2 Geometrical Information

All the features about the elements in the scene, and its relations are saved on matrix R, this information is used to construct the S matrix.

Matrix R is a three dimensional matrix with the information of each agent and each relationship presented at the scene.

A scene with N elements has $M = \frac{N*(N-1)}{2}$ possible edges that must be saved, and N elements features vectors.

Each vector of the matrix has P components (two or three for the geometrical information, and some more for the rest of the context information, depending on the problem domain).

The first ones N vectors represent the features of the N elements, and the next M vectors represent the features of the relationships between each element and other.

The R matrix has one row for each frame of the scene and N + M columns.

$$\begin{pmatrix} r_{11} & \cdots & r_{N1} & r_{(12)1} & \cdots & r_{(N-1N)1} \\ \vdots & \ddots & & & \vdots & \\ r_{1T} & \cdots & r_{NT} & r_{(12)T} & \cdots & r_{(N-1N)T} \end{pmatrix}$$

4.3 Semantics Information

The semantics information represents the associations between the elements of the scene to perform groups. One element could be part of many groups, and could be many groups.

This information makes it possible to create different associations between elements to grasp better the semantics context.

This semantics information is saved in a binary matrix with one row for each group, and one column for each element.

The matrix can only contain zeros or ones, which represent if this element forms part of the graph.

For example, in a scene with nine elements, and two groups, the matrix A could be like this one: $A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$. This matrix shows that there are two graphs, the first one composed by the elements: 1, 2, 3, 4, and 5; and the second one composed by the elements: 5, 6, 7, 8 and 9.

4.4 Structure Information

Matrices S define the structure of the graphs, there are one matrix for each graph.

Each S matrix has a number of T rows, and $M_g + N_g$ columns, where N_g is the number of elements of the group and M_g depends on N_g ($M_g = \frac{N_g*(N_g-1)}{2}$).

Each element of S is a features vector, describe in the section above. The selection of the important edges is made by the geometrical information of the features vector.

Each S matrix has the edges of the graph that defines its structure. If some edge have the same direction that another one and it is longer than the previous one, then this edge is not added to the matrix. One null value is added in this position.

Figure 4 shows the construction process: the shortest edge of the first element is added in (a). Then the second shortest is also added in (b). In (c) there is a shorter edge with the same direction (2) and the edge is not added. The process is repeated until all the elements are checked (d), (e) and (f).

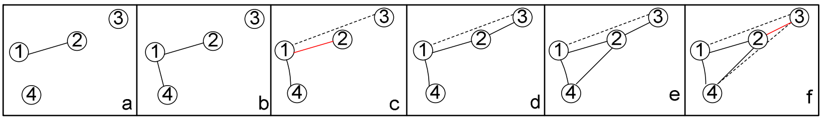

Fig. 3. Construction process

The matrix below shows the S of the graph in the Fig. 5b. First row represents the graph at the instant t = 0, and last row shows the graph at the instant t = T.

Fig. 5b shows an example of a graph with five elements where are presented the first two features (geometrical information). The first five columns represent the position of the elements, and the other represents the relationships. In the first frame the edges between the nodes 1-5, 3-4 and 3-5 are not defined because they have the same directions (and they are longer) than the edges 1-5, 3-2 and 5-2.

Then, in the frame T the graph's shape has changed, there are new relevant edges (like 3-5 and 3-4) and some relative distance have also changed.

$$\begin{pmatrix} 7^6 & 6^6 & 8^6 & 2^6 & 1^6 & 4^6 & 6^8 & 3^4 & -1^5 & 3^2 & 4^1 & 5^2 & -1^1 & -1^2 & 4^5 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 7^6 & 6^6 & 8^6 & 6^6 & 5^6 & 4^6 & 5^8 & 4^4 & -1^3 & 4^3 & 6^8 & 5^2 & 6^1 & 7^8 & 5^3 \end{pmatrix}$$

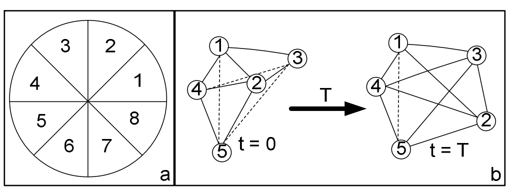

Fig. 4. a) Directions Code b) Example graph

5 Conclusions

In most of the domains of context-aware the users are not along and their behavior depends on the rest of the nearby users. Group behavior recognition could take advantage of the context information and it could be a good way to understand (and predict) the behaviors of the user on a pervasive computing system.

The structure of the information in context-aware systems is one of the most interesting fields of research, and it could be merging with the group behavior recognition systems.

Our approach had reduced the number of relations without loss information about the formation to realize the rezoning process. This approach is based in a novel structured representation of the important relations between the elements of the graphs.

The features selected in each case will depend on the problem domain, but in the most of the case the positioning system is available and provides a good source of knowledge.

Acknowledgements. This work was supported in part by Projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC2008-06732-C02-02/TEC, CAM CONTEXTS (S2009/ TIC-1485) and DPS2008-07029-C02-02.

References

1. Mäntyjärvi, Jani, Himberg, Johan and Huuskonen, Pertti: Collaborative Context Recognition for Handheld Device. Oulu : IEEE, (2003). Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on. pp. 161--168. 0769518931.
2. Mäntyjärvi, Jani and Seppänen: Adapting applications in handheld devices using fuzzy context information. Tapio. 4, Oulu : Elsevier, (2003), Interacting with Computers, Vol. 15, pp. 521--538. 0953-5438.
3. Baldauf, Matthias, Schahram, Dustdar and Rosenberg: A survey on context-aware systems. Florian. (2007), International Journal of Ad Hoc and Ubiquitous Computing, pp. 263--277.
4. Moeslund, T B, Kruger, V and Hilton: A survey of advances in vision-based human motion capture and analysis. A. 2-3, s.l. : Elsevier, (2006), Computer vision and image understanding, Vol. 104, pp. 90--126. 1077-3142.
5. Ryoo, M. S. and Aggarwal, J. K: Recognition of High-level Group Activities Based on Activities of Individual Members. s.l. : IEEE, (2008). Motion and video Computing, 2008. WMVC 2008. IEEE Workshop on. pp. 1--8.
6. Khan, Saad M. and Shah, Mubarak. Singapore: Detecting Group Activities using Rigidity of Formation. ACM, (2005). Proceedings of the 13th annual ACM international conference on Multimedia. pp. 403--406. 1595930442.
7. Ruonan, Li, Rama, Chellappa and Shaohua, Kevin Zhou: Learning Multi-modal Densities on Discriminative Temporal Interaction Manifold for Group Activity Recognition. New York : s.n., (2009), CVPR, pp. 1--8.
8. Perse, Matej, et al. 5, Ljubljana: A trajectory-based analysis of coordinated team activity in a basketball game. Elsevier, (2009), Computer Vision and Image Understanding, Vol. 113, pp. 612--621. 1077-3142.
9. Ramos, Fernando and Ayanegui, Huberto. Tracking behaviours of cooperative robots within multi-agent domains. Autonomous Agents. Tlaxcala : Vedran Kordic, (2010).